Nucleic Acids Research, 2007, Vol. 35, Web Server issue ***W305–W309***
doi:10.1093/nar/gkm255

RNAbor: a web server for RNA structural neighbors

Eva Freyhult[1], Vincent Moulton[2] and Peter Clote[3,*]

[1]Linnaeus Centre for Bioinformatics, Uppsala University, 75124 Uppsala, Sweden, [2]School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK and [3]Department of Biology, Boston College, Chestnut Hill, MA 02467, USA

Received January 24, 2007; Revised March 26, 2007; Accepted April 8, 2007

ABSTRACT

`RNAbor` provides a new tool for researchers in the biological and related sciences to explore important aspects of RNA secondary structure and folding pathways. `RNAbor` computes statistics concerning δ-neighbors of a given input RNA sequence and structure (the structure can, for example, be the minimum free energy (MFE) structure). A δ-neighbor is a structure that differs from the input structure by exactly δ base pairs, that is, it can be obtained from the input structure by adding and/or removing exactly δ base pairs. For each distance δ `RNAbor` computes the density of δ-neighbors, the number of δ-neighbors, and the MFE structure, or MFE^{δ} structure, among all δ-neighbors. `RNAbor` can be used to study possible folding pathways, to determine alternate low-energy structures, to predict potential nucleation sites and to explore structural neighbors of an intermediate, biologically active structure. The web server is available at `http://bioinformatics.bc.edu/clotelab/RNAbor`.

INTRODUCTION

RNA plays a surprising and previously unsuspected role in many biological processes, such as post-transcriptional regulation, conformational switches, expansion of the genetic code (such as selenocysteine insertion), ribosomal frameshift, metabolite-binding and chemical modification of specific nucleotides in the ribosome. Apart from its catalytic role as a ribonucleic enzyme (*ribozyme*) (1), RNA can regulate genes in several ways. For example, by hybridizing to a portion of messenger RNA, small $\sim$22 nt RNA molecules perform post-transcriptional gene regulation by RNA interference (RNAi), a process so important that for its discovery the 2006 Nobel Prize in Physiology or Medicine was awarded to A. Z. Fire and C. C. Mello. In addition, by very different means, RNA can perform transcriptional and translational gene regulation by allostery, where a portion of the 5′ untranslated region (5′ UTR) of mRNA, known as a *riboswitch* (2,3), undergoes a conformational change upon binding a specific ligand such as adenine, guanine or lysine.

As the field of *RNomics* matures, many sophisticated computational tools, e.g. RNA structure prediction, alignment and gene finding, have been developed—see (4,5) for recent overviews. Recently developed programs that are of most relevance here include the program `Sfold` (6,7), that computes a low energy ensemble of structures by sampling from the partition function (8), and an earlier program `RNAsubopt` (9) that computes all suboptimal structures within a user-specified number of kcal/mol of the minimum free energy (MFE). In addition, the program `RNAshapes` (10–12) provides a useful description of RNA branching structure by computing the Boltzmann probability of various shapes and also the MFE structure for various shapes. Here, an RNA shape is an equivalence class of secondary structures, describing the overall branching; for instance the shape of a typical cloverleaf tRNA would be [[][][]].

In this article, we describe the web server `RNAbor`, which computes the Boltzmann probability and MFE structures which differ by δ base pairs from a given initial structure. Unlike most of the tools just described, which focus on the MFE structure or a low energy ensemble, `RNAbor` yields information concerning the secondary structure folding landscape. Potential applications of `RNAbor` include the design of RNA aptamers (see (13) for a suggestion how RNA might be designed to inhibit the function of the viral enzymes such as HIV-1 reverse transcriptase and hepatitis C NS3 protease), detection of conformational switches, understanding the role played by biologically active structural intermediates and improvement in secondary structure prediction.

MATERIALS AND METHODS

Let $\mathbf{s}$ denote a given RNA nucleotide sequence, and let $\mathcal{S}$ be any given secondary structure of $\mathbf{s}$. The structure $\mathcal{S}$ could be the MFE structure of $\mathbf{s}$, it could be the secondary structure obtained from the 3-dimensional X-ray conformation or by comparative sequence analysis, or it could

*To whom correspondence should be addressed. Tel: +1 617 552 1332; Fax: +1 617 552 2011; Email: clote@bc.edu

be an arbitrary intermediate structure of particular biological significance. For an integer δ, a secondary structure $\mathcal{T}$ of $\mathbf{s}$ is a δ-neighbor of $\mathcal{S}$, if $\mathcal{S}$ and $\mathcal{T}$ differ by exactly δ base pairs [14]. In (Freyhult, E., Moulton, V. and Clote, P. Boltzmann probability of RNA structural neighbors and riboswitch detection, submitted for publication), we describe new algorithms, which compute the number $\mathbf{N}^\delta$ of δ-neighbors, the partition function $\mathbf{Z}^\delta$ for δ-neighbors and the MFE$^\delta$, and the corresponding MFE$^\delta$ structure over all δ-neighbors of a fixed structure $\mathcal{S}$.

Computing structural neighbors

To give the reader a feeling for how the algorithms work, we present the recurrence relations to compute the number $\mathbf{N}^\delta$ of δ-neighbors of $\mathcal{S}$. Let $\mathbf{s} = s_1, \ldots, s_n$. If $\mathbf{N}^\delta_{i,j}$ denotes the number of δ-neighbors of the substructure $\mathcal{S}_{[i,j]}$, the restriction of $\mathcal{S}$ to interval $[i,j]$ of $\mathbf{s}$, then the number of δ-neighbors of $\mathcal{S}$, $\mathbf{N}^\delta = \mathbf{N}^\delta_{1,n}$, can be computed by the following recursion:

$$\mathbf{N}^\delta_{i,j} = \mathbf{N}^{\delta-b_0}_{i,j-1} + \sum_{\substack{s_k s_j \in \mathbb{B},\\ i \leqslant k < j}} \sum_{w+w'=\delta-b} \mathbf{N}^{w}_{i,k-1}\mathbf{N}^{w'}_{k+1,j-1}, \qquad \mathbf{1}$$

where $\mathbb{B} = \{\mathrm{AU}, \mathrm{UA}, \mathrm{GC}, \mathrm{CG}, \mathrm{GU}, \mathrm{UG}\}$ (i.e. the set of Watson-Crick base pairs together with wobbles), $b_0 = 1$ if j is base-paired in $\mathcal{S}_{[i,j]}$ and 0 otherwise and b is the base pair distance between $\mathcal{S}_{[i,j]}$ and a structure on the same interval $[i,j]$ where a base pair between k and j has been added (taking into account all the base pairs in $\mathcal{S}_{[i,j]}$ that need to be broken to allow the addition of this base pair).

This approach for computing $\mathbf{N}^\delta$ can be extended to compute the partition function contribution, $\mathbf{Z}^\delta$, of the set of δ-neighbors and also to compute the MFE$^\delta$ and the MFE$^\delta$ structure. Computations are made with respect to the Turner energy model (15,16); treatment of the dangle is similar to that in Vienna RNA Package (option `-d2`). The algorithms employ dynamic programming, and run in $O(\Delta \cdot n^3)$ time and $O(\Delta \cdot n^2)$ space, where n is the sequence length and Δ is the maximum value of δ. Since Δ can be at most n, the run time cannot be worse than $O(n^4)$ and space no worse that $O(n^3)$, even if the user does not specify a value of Δ. Full details of the algorithms are given in (Freyhult, E., Moulton, V. and Clote, P. Boltzmann probability of RNA structural neighbors and riboswitch detection, submitted for publication).

Web server

The web server available at `http://bioinformatics.bc.edu/clotelab/RNAbor` runs on a Linux cluster with 20 computational nodes, each with double processors of between 1300 and 3000 MHz and 2 GB RAM (6 Dell PowerEdge 1650, 2 × 1300 MHz Pentium III, 2 GB RAM; 11 Dell PowerEdge 1850, 2 × 2800 + MHz Xeon EM64T, 2 GB RAM; 5 Dell PowerEdge 1850, 2 × 3000 MHz Xeon EM64T, 2 GB RAM).

RESULTS

Due to the time and space constraints of the algorithm, RNA sequences may be of length up to 300 nucleotides. Sequences of length up to 60 are processed interactively and output is displayed in the user's browser window. For sequences of length 61–300, the computation is done off-line and the results are returned to the user by email; for this, the email address is required. The user can either paste an input sequence (with optional secondary structure), or upload a file of the same. The full input consists of up to four lines, illustrated by the following example.

```
> 3UTR_MUSGBPA
AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU
((.((((..((((.............)))).)))).)).....
10
```

The temperature is set to a default value of 37°C; however the user can enter any integer temperature between 0 and 100.

The only required input is an RNA sequence $\mathbf{s}$ of length at most 300 nucleotides; the FASTA comment, initial secondary structure $\mathcal{S}$ and upper bound Δ are optional inputs. If no secondary structure is given, then the initial structure $\mathcal{S}$ is taken to be the MFE structure, as computed by `RNAfold -d2`. If the optional input Δ is missing, then Δ is defined to equal the length n of the input sequence $\mathbf{s}$; otherwise Δ is the minimum of the input value and n. For each $0 \leqslant \delta \leqslant \Delta$, `RNAbor` computes the Boltzmann probability $p^\delta = \mathbf{Z}^\delta/\mathbf{Z}$, where the partition function is defined by

$$\mathbf{Z}^\delta = \sum_{\mathcal{T}} e^{-E(\mathcal{T})/RT},$$

where R is the universal gas constant and T is temperature in degrees Kelvin. Here, the summation is made over all secondary structures $\mathcal{T}$ of $\mathbf{s}$ which are δ-neighbors of $\mathcal{S}$. The full partition function $\mathbf{Z} = \sum_\delta \mathbf{Z}^\delta$ is computed by McCaskill's algorithm (8) if $\Delta \geqslant n$.

In addition to computing probability p^δ, `RNAbor` computes the number N^δ of δ-neighbors of $\mathcal{S}$, the MFE$^\delta$ over all δ-neighbors of $\mathcal{S}$ and the MFE$^\delta$ secondary structure. Tables of the values N^δ and p^δ, as well as their graphs, are made available as downloadable files. The five-column text file output, consisting of δ, p^δ, N^δ, MFE$^\delta$ and the MFE$^\delta$ structure, is depicted in Figure 1.

EXAMPLES

`RNAbor` can be used to generate alternative low energy structures, which differ markedly from the MFE structure, or from any initially given structure. Figure 1 shows the `RNAbor` output for a short 3′-UTR sequence of an mRNA with NCBI accession number MUSGBPS. The input structure in this example is the MFE structure (as predicted by `RNAfold -d2`). The `RNAbor` output indicates two ranges of δ that show higher probabilities than the rest, 0–9 and 20–24. The MFE$^\delta$ structures at

RNAbor - structural neighbors for RNA secondary structure

> 3UTR_MUSGBPA

RNA sequence: AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU

Secondary structure: ((.((((..((((.....))))).)))).))....

Temperature: 37

Output of Boltzmann probability of secondary structures at base pair distance k from minimum free energy structure (i.e. k-neighbors), the corresponding graph, as well as number of k-neighbors and corresponding graph.

- Boltzmann probabilities
- Graph of Boltzmann probabilities
- Number of k-neighbors
- Graph of number of k-neighbors
- Entire text output of RNAbor

```
#RNAbor kneighbor entire output for rna  3UTR_MUSGBPA
# AGCCAGCCAGCCUGUAGCCCUCAAUAAAAGGCAGCUGCCUCUGCUCCCCAU
# ((.((((..((((.....))))).)))).))....
```

k	Density	NumStr	Energy	MFE(k)
0	0.098606	1	10.060000	((.((((..((((.....))))).)))).))....
1	0.081764	35	9.760000	.(.((((..((((.....))))).)))).).....
2	0.081878	471	9.460000	...((((..((((.....))))).)))).....
3	0.070206	3664	9.360000	.(.((((..((((.....))))).)))).....
4	0.077459	19885	9.560000	.(((((..((((.....))))).))))...))....
5	0.130056	83810	9.860000	(((((((..((((.....))))).))))...)))....
6	0.041013	293123	8.700000	.(((((..((((.((.....)).)))).))))...))....
7	0.036915	887181	9.000000	(((((((..((((.((.....)).)))).))))...)))....
8	0.010103	2386686	8.200000	(((((((..(((..((.....)).)))).))))...)))....
9	0.003034	5820155	6.900000	(((((((..((((((.....))).)))).))))...)))....
10	0.000817	13132239	5.800000	(((((((..((((((.(....).))).)))).))))...)))....
11	0.000537	28037969	5.800000	.........((((.((.....)).)))).(((.....)))....
12	0.000406	57646218	5.160000	.((.((.((((((.............))))...)).)).))....
13	0.000407	115302366	5.460000	(((.((.((((((.............))))...)).)).)))...
14	0.000284	230402595	5.300000	(((((((..((.....))(((.....))).)))).))...)))....
15	0.000217	497103424	4.800000	..........................((.(((.....))).))....
16	0.000301	1147035433	4.900000	.((.....)).................((((....))))....
17	0.001049	2260415939	5.700000	.........((..((.(((.....))).)).))....
18	0.003151	3076558005	6.700000	............(((((((.....))).))))....
19	0.007917	2634229251	6.700000	.((....(((.....(((.....))).)))).))...))....
20	0.025394	1377778726	7.700000	.........((..(((((((.....))).)))).))....))....
21	0.063371	436892492	8.500000	.......(((...(((((((.....))).)))).))...)))....
22	0.103977	84071172	8.800000	.....((.((...(((((((.....))).)))).))...))))....
23	0.110059	9946306	9.400000	....(((.((...(((((((.....))).)))).))...))))....
24	0.049669	728473	8.600000	.(..(((.((...(((((((.....))).)))).))...)))..)...
25	0.001410	27437	7.000000	(((.((..((...(((((((.....))).)))).))...)))....
26	0.000000	214	3.100000	(((.((..((...(((((((.....))).)))).))))))))....))
27	0.000000	0	0.000000	
28	0.000000	0	0.000000	
29	0.000000	0	0.000000	
30	0.000000	0	0.000000	
31	0.000000	0	0.000000	

Figure 1. Text output of `RNAbor` on the 51 nt 3′ UTR of a mRNA with NCBI accession number MUSGBPS. The five columns in the entire text output from `RNAbor` are given by the following, in order: (i) value of δ, (ii) Boltzmann probability p^{δ}, (iii) number of δ-neighbors, (iv) MFE of δ-neighbors, denoted MFE$^{\delta}$, (v) MFE secondary structure among all δ-neighbors. In this case, the initial structure is the MFE structure, as determined by `RNAfold -d2`.

distance δ between 0 and 9 from the MFE structure all have very similar folds and the probability of finding the RNA in a structure at δ between 0 and 9 is 0.63. The probability of finding a structure at δ 20–24 is also relatively high, 0.35, and the MFE$^{\delta}$ structures in this range are similar to each other but completely different from the MFE structure. Thus the two highly probable δ ranges represent two possible alternative folds of the RNA.

Analyzing the same sequence with `Sfold` gives similar results. `Sfold` finds three types of structures (three clusters), with probabilities 0.65, 0.22 and 0.13, respectively. One cluster contains the MFE structure

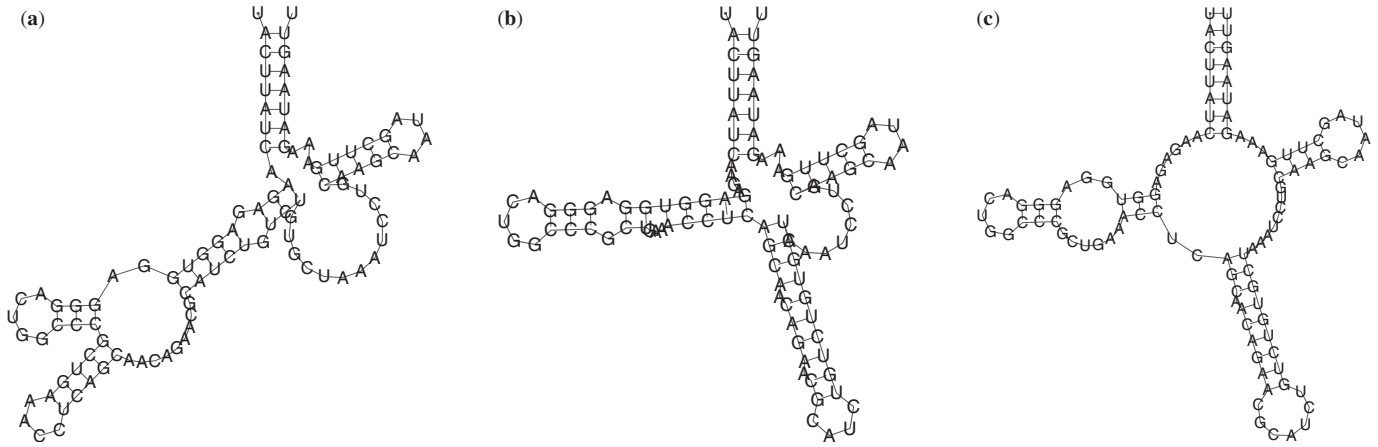

Figure 2. Two alternative low energy secondary structures for the 101 nt SAM riboswitch with EMBL accession number AP004597.1 from position 118941 to position 119041. This riboswitch has nucleotide sequence UACUUAUCAA GAGAGGUGGA GGGACUGGCC CGCUGAAACC UCAGCAACAG AACGCAUCUG UCUGUGCUAA AUCCUGCAAG CAAUAGCUUG AAAGAUAAGU U. Panel (**a**) displays the MFE structure with free energy −33.30 kcal/mol, while (**b**) displays the 30-neighbor of the MFE with free energy of −32.10 kcal/mol. The only significant Boltzmann probabilities were for values around $\delta = 0$ and $\delta = 30$, where $p^0 = 0.056238$ and $p^{30} = 0.151751$. Note that the MFE^{30} structure more closely resembles the expected structure for a riboswitch shown in (**c**), as determined from the Rfam (17) consensus structure.

corresponding to the folds at δ values from 0 to 9, another cluster has a centroid structure resembling the structures at δ between 20 and 24, and the third cluster has a centroid structure similar to the MFE[19] structure. `RNAshapes` on the other hand is less successful for this example since the alternative folds as predicted by `RNAbor` have the same shape [ ], even though the folds are very different.

Figure 2 displays the MFE structure and the MFE^{30} structure of the 101 nt SAM riboswitch with EMBL accession number AP004597.1/118941-119041, with sequence taken from Rfam (17). The MFE structure over all 30-neighbors, the MFE^{30} structure, is clearly much closer to the real structure than the global MFE structure. Figure 3 displays the Boltzmann probability density, showing a peak for the value $\delta = 30$.

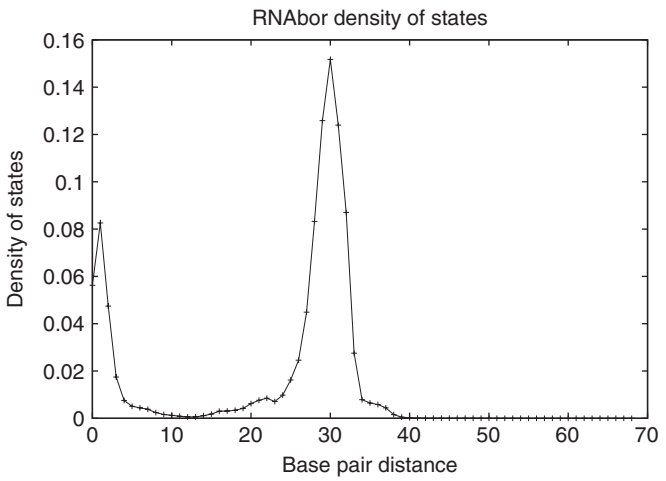

Figure 3. Boltzmann probability density plot for the 101 nt SAM riboswitch (EMBL accession number AP004597.1/118941-119041). The curve shows the probability, $p^\delta = \mathbf{Z}^\delta/\mathbf{Z}$, for all secondary structures of RNA sequence **s** having base pair distance δ from the MFE structure $\mathcal{S}$.

DISCUSSION

In this article, we have introduced the web server `RNAbor`, which computes the Boltzmann probability and MFE structure over all δ-neighbors for a given RNA sequence **s** and initial secondary structure $\mathcal{S}$. The underlying algorithms, described in the forthcoming paper (Freyhult, E., Moulton, V. and Clote, P. Boltzmann probability of RNA structural neighbors and riboswitch detection, submitted for publication), use dynamic programming, involve the Turner energy model (15,16), and require considerable time $O(\Delta \cdot n^3)$ and space $O(\Delta \cdot n^2)$ resources. Figures 2 and 3 illustrate the use of `RNAbor` in better understanding structural aspects of a SAM riboswitch, and indicate that `RNAbor` should provide a useful complementary tool to programs such as `Sfold` and `RNAshapes` for analyzing the ensemble of possible secondary structures on a given RNA sequence.

ACKNOWLEDGEMENTS

Research of P.C. was partially supported by National Science Foundation DBI-0543506, which additionally supported some travel of E.F. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. All three authors would like to thank Elena Rivas, Eric Westhof and funding agencies for organizing the meeting RNA-2006 in Benasque, Spain, in July 2006, where some of this work was carried out. Finally, thanks to Jason Persampieri for some technical assistance. Funding to pay

the Open Access publication charges for this article was provided by the National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Doudna,J.A. and Cech,T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222–228.
2. Winkler,W.C., Cohen-Chalamish,S. and Breaker,R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad. Sci. USA*, **99**, 15908–15913.
3. Penchovsky,R. and Breaker,R.R. (2005) Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nat. Biotechnol.*, **23**, 1424–1431.
4. Mathews,D.H. and Turner,D.H. (2006) Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.*, **16**, 270–278.
5. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
6. Ding,Y. and Lawrence,C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.
7. Ding,Y., Chan,C.Y. and Lawrence,C.E. (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, **11**, 1157–1166.
8. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structures. *Biopolymers*, **29**, 1105–1119.
9. Wuchty,S., Fontana,W., Hofacker,I.L. and Schuster,P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–164.
10. Giegerich,R., Voss,B. and Rehmsmeier,M. (2004) Abstract shapes of RNA. *Nucleic Acids Res.*, **32**, 4843–4851.
11. Steffen,P., Voss,B., Rehmsmeier,M., Reeder,J. and Giegerich,R. (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, **22**, 500–503.
12. Voss,B., Giegerich,R. and Rehmsmeier,M. (2006) Complete probabilistic analysis of RNA shapes. *BMC Biol.*, **4**, 5.
13. James,W. (2001) Nucleic acid and polypeptide aptamers: a powerful approach to ligand discovery. *Curr. Opin. Pharmacol.*, **1**, 540–548.
14. Moulton,V., Zuker,M., Steel,M., Pointon,R. and Penny,D. (2000) Metrics on RNA secondary structures. *J. Comput. Biol.*, **7**, 277–292.
15. Matthews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
16. Xia,T., SantaLucia,J.Jr, Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.
17. Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A. and Eddy,S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.